PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C12P 21/02, C12N 15/11, 15/33, 15/48, 15/85 — A1

(11) International Publication Number: WO 98/12207

(43) International Publication Date: 26 March 1998 (26.03.98)

(21) International Application Number: PCT/US97/16639

(22) International Filing Date: 18 September 1997 (18.09.97)

(30) Priority Data:
08/717,294 20 September 1996 (20.09.96) US

(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US).

(72) Inventors: SEED, Brian; Apartment 5J, Nine Hawthorne Place, Boston, MA 02114 (US). HAAS, Jurgen; Huberweg 13, D-69198 Schriesheim (DE).

(74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: HIGH LEVEL EXPRESSION OF PROTEINS

(57) Abstract

The invention features a synthetic gene encoding a protein normally expressed in a mammalian cell wherein at least one non-preferred or less preferred codon in the natural gene encoding the protein has been replaced by a preferred codon encoding the same amino acid.

	Leu	CTG	118.00	81.21	0.94
	Leu	CTA	3.00	2.06	0.02
	Leu	CTT	1.00	0.69	0.01
5	Leu	CTC	3.00	2.06	0.02
	Pro	CCG	4.00	2.75	0.05
	Pro	CCA	0.00	0.00	0.00
	Pro	CCT	3.00	2.06	0.04
10	Pro	CCC	68.00	46.80	0.91

TABLE 4: Codon Frequency Table of the Native Factor VIII B Domain Deleted Gene

AA	Codon	Number	/1000	Fraction
Gly	GGG	12.00	8.26	0.15
Gly	GGA	34.00	23.40	0.41
Gly	GGT	16.00	11.01	0.20
Gly	GGC	20.00	13.76	0.24
Glu	GAG	33.00	22.71	0.39
Glu	GAA	51.00	35.10	0.61
Asp	GAT	55.00	37.85	0.67
Asp	GAC	27.00	18.58	0.33
Val	GTG	29.00	19.96	0.33
Val	GTA	19.00	13.08	0.22
Val	GTT	17.00	11.70	0.19
Val	GTC	23.00	15.83	0.26
Ala	GCG	2.00	1.38	0.03
Ala	GCA	18.00	12.39	0.25
Ala	GCT	31.00	21.34	0.44
Ala	GCC	20.00	13.76	0.28

Arg	AGG	18.00	12.39	0.25
Arg	AGA	22.00	15.14	0.30
Ser	AGT	22.00	15.14	0.18
Ser	AGC	24.00	16.52	0.20
Lys	AAG	32.00	22.02	0.40
Lys	AAA	48.00	33.04	0.60
Asn	AAT	38.00	26.15	0.60
Asn	AAC	25.00	17.21	0.40
Met	ATG	43.00	29.59	1.00
Ile	ATA	13.00	8.95	0.18
Ile	ATT	36.00	24.78	0.49
Ile	ATC	25.00	17.21	0.34
Thr	ACG	1.00	0.69	0.01
Thr	ACA	23.00	15.83	0.28
Thr	ACT	36.00	24.78	0.43
Thr	ACC	23.00	15.83	0.28
Trp	TGG	28.00	19.27	1.00
End	TGA	1.00	0.69	1.00
Cys	TGT	7.00	4.82	0.37
Cys	TGC	12.00	8.26	0.63
End	TAG	0.00	0.00	0.00
End	TAA	0.00	0.00	0.00
Tyr	TAT	41.00	28.22	0.60
Tyr	TAC	27.00	18.58	0.40
Leu	TTG	20.00	13.76	0.16
Leu	TTA	10.00	6.88	0.08
Phe	TTT	45.00	30.97	0.58
Phe	TTC	32.00	22.02	0.42
Ser	TCG	2.00	1.38	0.02
Ser	TCA	27.00	18.58	0.22
Ser	TCT	27.00	18.58	0.22
Ser	TCC	18.00	12.39	0.15

Arg	CGG	6.00	4.13	0.08
Arg	CGA	10.00	6.88	0.14
Arg	CGT	7.00	4.82	0.10
Arg	CGC	10.00	6.88	0.14
Gln	CAG	42.00	28.91	0.63
Gln	CAA	25.00	17.21	0.37
His	CAT	28.00	19.27	0.55
His	CAC	23.00	15.83	0.45
Leu	CTG	36.00	24.78	0.29
Leu	CTA	15.00	10.32	0.12
Leu	CTT	24.00	16.52	0.19
Leu	CTC	20.00	13.76	0.16
Pro	CCG	1.00	0.69	0.01
Pro	CCA	32.00	22.02	0.43
Pro	CCT	26.00	17.89	0.35
Pro	CCC	15.00	10.32	0.20

Use

The synthetic genes of the invention are useful for expressing the a protein normally expressed in mammalian cells in cell culture (e.g. for commercial production of human proteins such as hGH, TPA, Factor VIII, and Factor IX). The synthetic genes of the invention are also useful for gene therapy. For example, a synthetic gene encoding a selected protein can be introduced in to a cell which can express the protein to create a cell which can be administered to a patient in need of the protein. Such cell-based gene therapy techniques are well known to those skilled in the art, see, e.g., Anderson, et al., U.S. Patent No. 5,399,349; Mulligan and Wilson, U.S. Patent No. 5,460,959.

What is claimed is:

1. A synthetic gene encoding a protein normally expressed in an eukaryotic cell wherein at least one non-preferred or less preferred codon in a natural gene encoding said protein has been replaced by a preferred codon encoding the same amino acid, said synthetic gene being capable of expressing said protein at a level which is at least 110% of that expressed by said natural gene in an *in vitro* mammalian cell culture system under identical conditions.

2. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said protein at a level which is at least 150% of that expressed by said natural gene in an *in vitro* cell culture system under identical conditions.

3. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said protein at a level which is at least 200% of that expressed by said natural gene in an *in vitro* cell culture system under identical conditions.

4. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said protein at a level which is at least 500% of that expressed by said natural gene in an *in vitro* cell culture system under identical conditions.

5. The synthetic gene of claim 1 wherein said synthetic gene comprises fewer than 5 occurrences of the sequence CG.

6. The synthetic gene of claim 1 wherein at least 10% of the codons in said natural gene are non-preferred codons.

7. The synthetic gene of claim 1 wherein at least 50% of the codons in said natural gene are non-preferred codons.

8. The synthetic gene of claim 1 wherein at least 50% of the non-preferred codons and less preferred codons present in said natural gene have been replaced by preferred codons.

9. The synthetic gene of claim 1 wherein at least 90% of the non-preferred codons and less preferred codons present in said natural gene have been replaced by preferred codons.

10. The synthetic gene of claim 1 wherein said protein is normally expressed by a mammalian cell.

11. The synthetic gene of claim 1 wherein said protein is a retroviral protein.

12. The synthetic gene of claim 1 wherein said protein is a lentiviral protein.

13. The synthetic gene of claim 11 wherein said protein is an HIV protein.

14. The synthetic gene of claim 13 wherein said protein is selected from the group consisting of gag, pol, and env.

15. The synthetic gene of claim 13 wherein said protein is gp120.

16. The synthetic gene of claim 13 wherein said protein is gp160.

17. The synthetic gene of claim 1 wherein said protein is a human protein.

18. The synthetic gene of claim 1 wherein said human protein is Factor VIII.

19. The synthetic gene of claim 1 wherein 20% of the codons are preferred codons.

20. The synthetic gene of claim 18 wherein said gene has the coding sequence present in SEQ ID NO:42.

21. The synthetic gene of claim 1 wherein said protein is green fluorescent protein.

22. The synthetic gene of claim 20 wherein said synthetic gene is capable of expressing said green fluorescent protein at a level which is at least 200% of that expressed by said natural gene in an *in vitro* mammalian cell culture system under identical conditions.

23. The synthetic gene of claim 20 wherein said synthetic gene is capable of expressing said green fluorescent protein at a level which is at least 1000% of that expressed by said n atural gene in an *in vitro* mammalian cell culture system under identical conditions.

24. The synthetic gene of claim 21 having the sequence depicted in Figure 11 (SEQ ID NO:40).

25. An expression vector comprising the synthetic gene of claim 1.

26. The expression vector of claim 21, said expression vector being a mammalian expression vector.

27. A mammalian cell harboring with the synthetic gene of claim 1.

28. A method for preparing a synthetic gene encoding a protein normally expressed by mammalian cells, comprising identifying non-preferred and less-preferred codons in the natural gene encoding said protein and replacing one or more of said non-preferred and less-preferred codons with a preferred codon encoding the same amino acid as the replaced codon.

Syngp120mn 1/18

```
   1  CTCGAGATCC ATTGTGCTCT AAAGGAGATA CCCGGCCAGA CACCCTCACC
  51  TGCGGTGCCC AGCTGCCCAG GCTGAGGCAA GAGAAGGCCA GAAACCATGC
 101  CCATGGGGTC TCTGCAACCG CTGGCCACCT TGTACCTGCT GGGGATGCTG
 151  GTCGCTTCCG TGCTAGCCAC CGAGAAGCTG TGGGTGACCG TGTACTACGG
 201  CGTGCCCGTG TGGAAGGAGG CCACCACCAC CCTGTTCTGC GCCAGCGACG
 251  CCAAGGCGTA CGACACCGAG GTGCACAACG TGTGGGCCAC CCAGGCGTGC
 301  GTGCCCACCG ACCCCAACCC CCAGGAGGTG GAGCTCGTGA ACGTGACCGA
 351  GAACTTCAAC ATGTGGAAGA ACAACATGGT GGAGCAGATG CATGAGGACA
 401  TCATCAGCCT GTGGGACCAG AGCCTGAAGC CCTGCGTGAA GCTGACCCCC
 451  CTGTGCGTGA CCCTGAACTG CACCGACCTG AGGAACACCA CCAACACCAA
 501  CAACAGCACC GCCAACAACA ACAGCAACAG CGAGGGCACC ATCAAGGGCG
 551  GCGAGATGAA CAACTGCAGC TTCAACATCA CCACCAGCAT CCGCGACAAG
 601  ATGCAGAAGG AGTACGCCCT GCTGTACAAG CTGGATATCG TGAGCATCGA
 651  CAACGACAGC ACCAGCTACC GCCTGATCTC CTGCAACACC AGCGTGATCA
 701  CCCAGGCCTG CCCCAAGATC AGCTTCGAGC CCATCCCCAT CCACTACTGC
 751  GCCCCCGCCG GCTTCGCCAT CCTGAAGTGC AACGACAAGA AGTTCAGCGG
 801  CAAGGGCAGC TGCAAGAACG TGAGCACCGT GCAGTGCACC CACGGCATCC
 851  GGCCGGTGGT GAGCACCCAG CTCCTGCTGA ACGGCAGCCT GGCCGAGGAG
 901  GAGGTGGTGA TCCGCAGCGA GAACTTCACC GACAACGCCA AGACCATCAT
 951  CGTGCACCTG AATGAGAGCG TGCAGATCAA CTGCACGCGT CCCAACTACA
1001  ACAAGCGCAA GCGCATCCAC ATCGGCCCCG GGCGCGCCTT CTACACCACC
1051  AAGAACATCA TCGGCACCAT CCGCCAGGCC CACTGCAACA TCTCTAGAGC
1101  CAAGTGGAAC GACACCCTGC GCCAGATCGT GAGCAAGCTG AAGGAGCAGT
1151  TCAAGAACAA GACCATCGTG TTCAACCAGA GCAGCGGCGG CGACCCCGAG
1201  ATCGTGATGC ACAGCTTCAA CTGCGGCGGC GAATTCTTCT ACTGCAACAC
1251  CAGCCCCCTG TTCAACAGCA CCTGGAACGG CAACAACACC TGGAACAACA
1301  CCACCGGCAG CAACAACAAT ATTACCCTCC AGTGCAAGAT CAAGCAGATC
1351  ATCAACATGT GGCAGGAGGT GGGCAAGGCC ATGTACGCCC CCCCCATCGA
1401  GGGCCAGATC CGGTGCAGCA GCAACATCAC CGGTCTGCTG CTGACCCGCG
1451  ACGGCGGCAA GGACACCGAC ACCAACGACA CCGAAATCTT CCGCCCCGGC
```

FIG 1
(SHEET 1 OF 4)

1501 GGGGGCGACA TGCGCGACAA CTGGAGATCT GAGCTGTACA AGTACAAGGT

1551 GGTGACGATC GAGCCCCTGG GCGTGGCCCC CACCAAGGCC AAGCGCCGCG

1601 TGGTGCAGCG CGAGAAGCGC TAAAGCGGCC GC (SEQ ID NO:34)

FIG 1

(SHEET 2 OF 4)

SyngpI60mn

```
   1  ACCGAGAAGC TGTGGGTGAC CGTGTACTAC GGCGTGCCCG TGTGGAAGGA
  51  GGCCACCACC ACCCTGTTCT GCGCCAGCGA CGCCAAGGCG TACGACACCG
 101  AGGTGCACAA CGTGTGGGCC ACCCAGGCCT GCGTGCCCAC CGACCCCAAC
 151  CCCCAGGAGG TGGAGCTCGT GAACGTGACC GAGAACTTCA ACATGTGGAA
 201  GAACAACATG CTGGAGCAGA TGCATGAGGA CATCATCAGC CTGTGGGACC
 251  AGAGCCTGAA GCCCTGCGTG AAGCTGACCC CCCTGTGCGT GACCCTCAAC
 301  TGCACCGACC TGAGGAACAC CACCAACACC AACAACAGCA CCGCCAACAA
 351  CAACAGCAAC AGCGAGGGCA CCATCAAGGG CGGCGAGATG AAGAACTGCA
 401  GCTTCAACAT CACCACCAGC ATCCGCGACA AGATCCAGAA GGAGTACGCC
 451  CTGCTGTACA AGCTGGATAT CGTGAGCATC GACAACGACA GCACCAGCTA
 501  CCGCCTGATC TCCTGCAACA CCAGCGTGAT CACCCAGGCC TGCCCCAAGA
 551  TCAGCTTCGA GCCCATCCCC ATCCACTACT GCGCCCCCGC CGGCTTCGCC
 601  ATCCTGAAGT GCAACGACAA GAAGTTCAGC GGCAAGGGCA GCTGCAAGAA
 651  CGTGAGCACC GTGCAGTGCA CCCACGGCAT CCGGCCGGTG GTGAGCACCC
 701  ACCTCCTGCT GAACGGCAGC CTGGCCGAGG AGGAGGTGGT GATCCGCAGC
 751  GAGAACTTCA CCGACAACGC CAAGACCATC ATCGTGCACC TGAATGAGAG
 801  CGTGCAGATC AACTGCACGC GTCCCAACTA CAACAAGCGC AAGCGCATCC
 851  ACATCGGCCC CGGGCGCGCC TTCTACACCA CCAAGAACAT CATCGGCACC
 901  ATCCGCCAGG CCCACTGCAA CATCTCTAGA GCCAAGTGGA ACGACACCCT
 951  GCGCCAGATC GTGAGCAAGC TGAAGGAGCA GTTCAAGAAC AAGACCATCG
1001  TGTTCAACCA GAGCAGCGGC GGCGACCCCG AGATCGTGAT GCACAGCTTC
1051  AACTGCGGCG GCGAATTCTT CTACTGCAAC ACCAGCCCCC TGTTCAACAG
1101  CACCTGGAAC GGCAACAACA CCTGGAACAA CACCACCGGC AGCAACAACA
1151  ATATTACCCT CCAGTGCAAG ATCAAGCAGA TCATCAACAT GTGGCAGGAG
1201  GTGGGCAAGG CCATGTACGC CCCCCCCATC GAGGGCCAGA TCCGCTGCAG
1251  CAGCAACATC ACCGGTCTGC TGCTGACCCG CGACGGCGGC AACGACACCG
1301  ACACCAACGA CACCGAAATC TTCCGCCCCG GCGGCGGCGA CATGCGCGAC
1351  AACTGGAGAT CTGAGCTGTA CAAGTACAAG GTGGTGACGA TCGAGCCCCT
1401  GGGCGTGGCC CCCACCAAGG CCAAGCGCCG CGTGGTGCAG CGCGAGAAGC
```

FIG. 1
(SHEET 3 OF 4)

1451 GGGCCGCCAT CGGCGCCCTG TTCCTGGGCT TCCTGGGGGC GGCGGGCAGC

1501 ACCATGGGGG CCGCCAGCGT GACCCTGACC GTGCAGGCCC GCCTGCTCCT

1551 GAGCGGCATC GTGCAGCAGC AGAACAACCT CCTCCGCGCC ATCGAGGCCC

1601 AGCAGCATAT GCTCCAGCTC ACCGTGTGGG GCATCAAGCA GCTCCAGGCC

1651 CGCGTGCTGG CGTGGAGCG CTACCTGAAG GACCAGCAGC TCCTGGGCTT

1701 CTGGGGCTGC TCCGGCAAGC TGATCTGCAC CACCACGGTA CCCTGGAACG

1751 CCTCCTGGAG CAACAAGAGC CTGGACGACA TCTGGAACAA CATGACCTGG

1801 ATGCAGTGGG AGCGCGAGAT CGATAACTAC ACCAGCCTGA TCTACAGCCT

1851 GCTGGAGAAG AGCCAGACCC AGCAGGAGAA GAACGAGCAG GAGCTGCTGG

1901 AGCTGGACAA CTGGGCGAGC CTGTGGAACT GGTTCGACAT CACCAACTGG

1951 CTGTGGTACA TCAAAATCTT CATCATGATT GTGGGCGGCC TGGTGGGCCT

2001 CCGCATCGTG TTCGCCGTGC TGAGCATCGT GAACCGCGTG CGCCAGGGCT

2051 ACAGCCCCCT GAGCTTCCAG ACCCGGCCCC CCGTGCCGCG CGGGCCCGAC

2101 CGCCCCGAGG GCATCGAGGA GGAGGGCGGC GAGCGCGACC GCGACACCAG

2151 CGGCAGGCTC GTGCACGGCT TCCTGGCGAT CATCTGGGTC GACCTCCGCA

2201 GCCTGTTCCT GTTCAGCTAC CACCACCGCG ACCTGCTGCT GATCGCCGCC

2251 CGCATCGTGG AACTCCTAGG CCGCCGCGGC TGGGAGGTGC TGAAGTACTG

2301 GTGGAACCTC CTCCAGTATT GGAGCCAGGA GCTGAAGTCC AGCGCCGTGA

2351 GCCTGCTGAA CGCCACCGCC ATCGCCGTGG CCGAGGGCAC CGACCGCGTG

2401 ATCGAGGTGC TCCAGAGGGC CGGGAGGGCG ATCCTGCACA TCCCCACCCG

2451 CATCCGCCAG CGGCTCGAGA GGGCGCTGCT G (SEQ ID NO:35)

FIG. 1

(SHEET 4 OF 4)

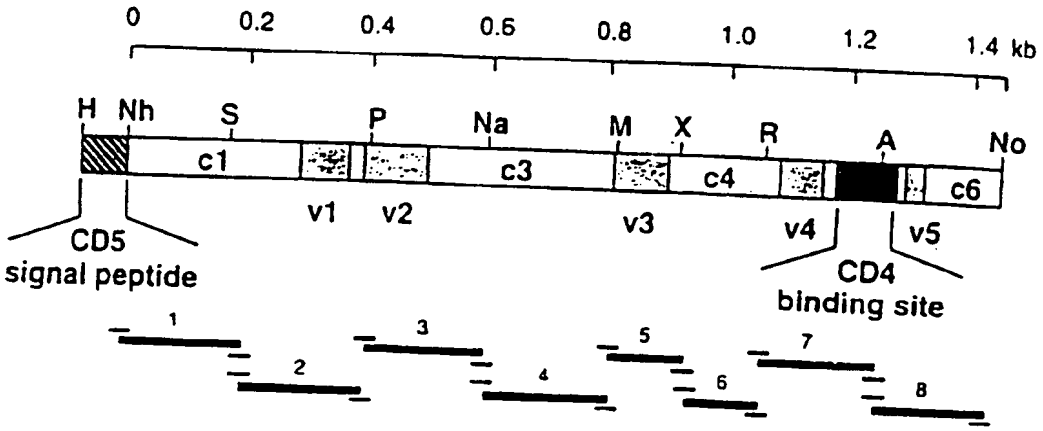

0
0.2
0.4
0.6
0.8
1.0
1.2
1.4 kb
H
Nh
S
P
Na
M
X
R
A
No
c1
c3
c4
c6
v1
v2
v3
v4
v5
CD5
signal peptide
CD4
binding site
1
2
3
4
5
6
7
8

FIGURE 2

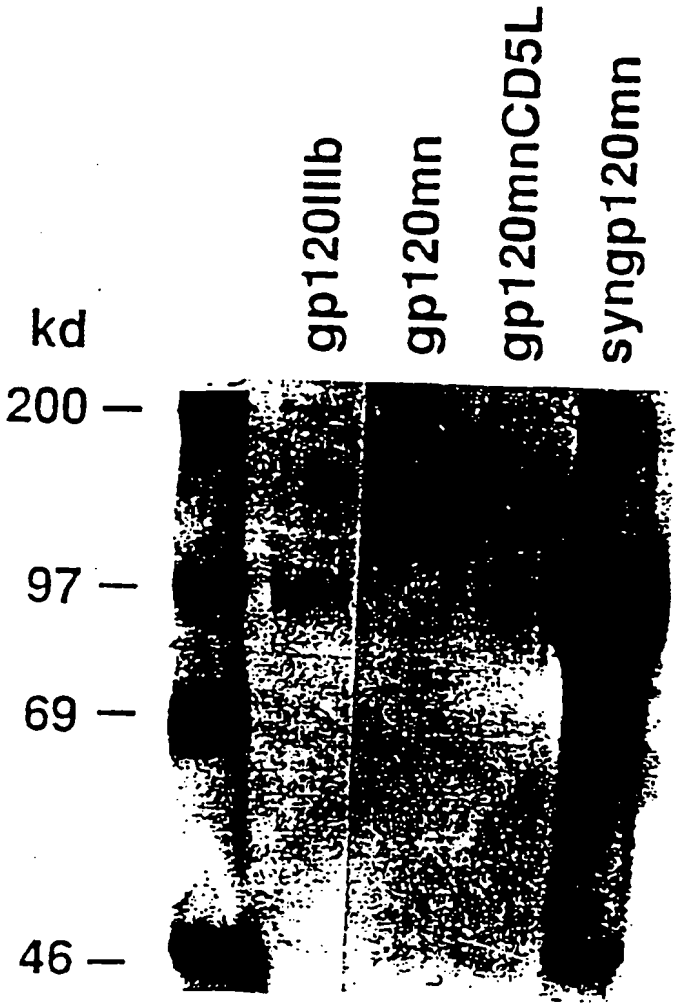


FIGURE 3

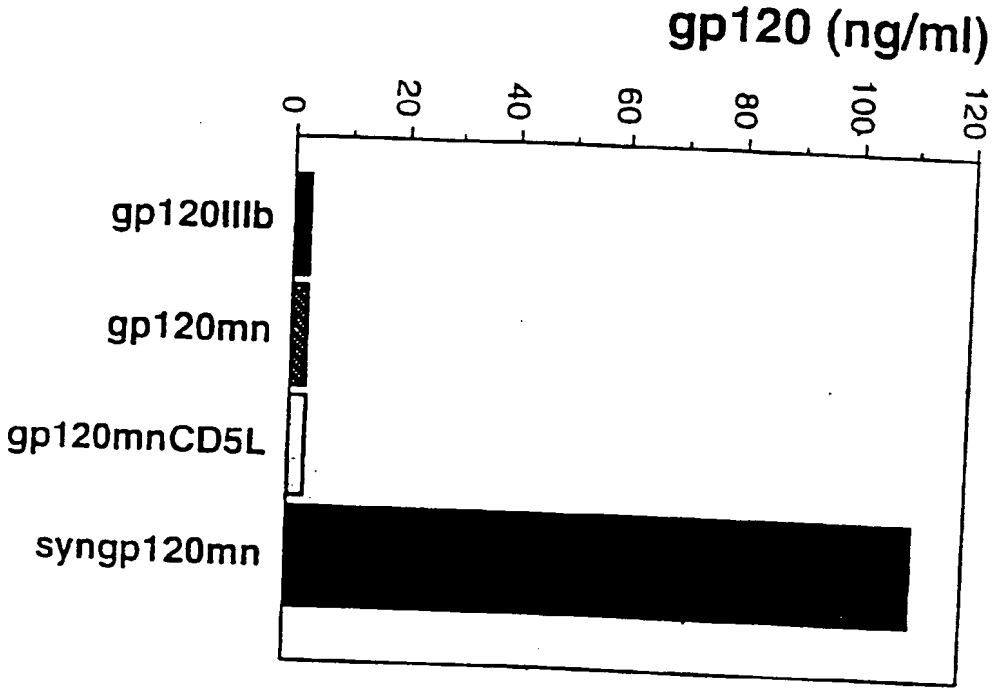

gp120 (ng/ml)
0
20
40
60
80
100
120
gp120IIIb
gp120mn
gp120mnCD5L
syngp120mn

FIGURE 4

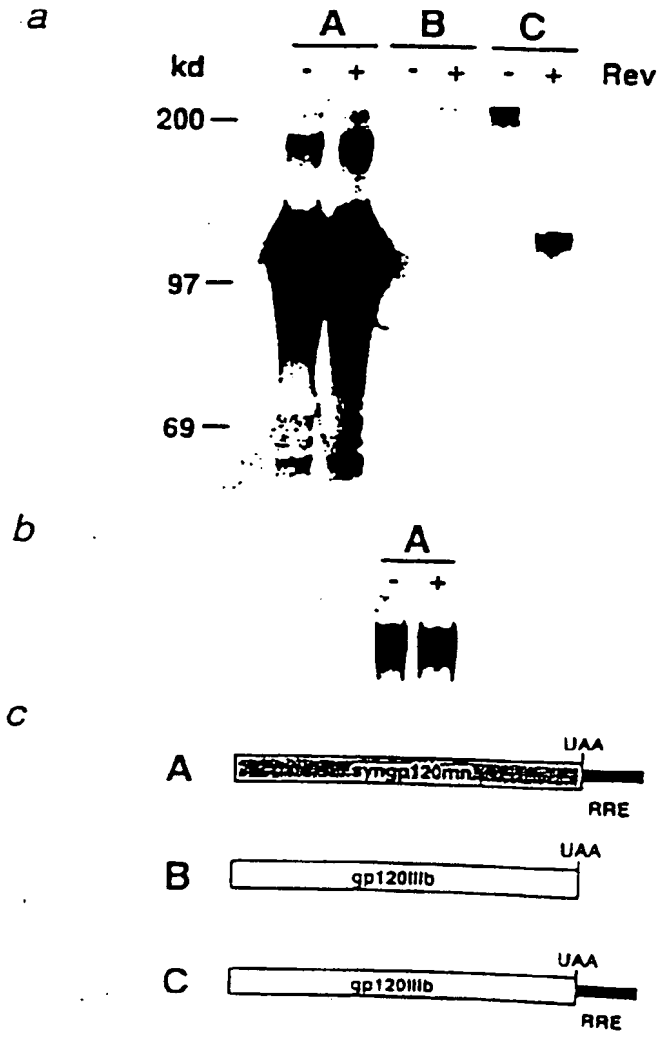

a
A
B
C
kd
- + - + - + Rev
200
97
69
b
A
- +
c
A
UAA
RRE
B
UAA
gp120IIIb
C
UAA
gp120IIIb
RRE

FIGURE 5

```
                  M   N   P   V   I   S   I   T   L   L   L   S   V   L   Q   M   S   R   G   Q
(SEQ ID NO:36) env→atg aat cca gta ata agt ata aca tta tta tta agt gta tta caa atg agt aga gga caa
(SEQ ID NO:37) wt→ atg aac cca gtc atc agc atc act ctc ctg ctt tca gtc ttg cag atg tcc cga gga cag

                   R   V   I   S   L   T   A   C   L   V   N   Q   N   L   R   L   D   C   R   H
env     aga gta ata agt tta aca gca tgt tta gta aat caa aat ttg aga tta gat tgt aga cat
wt      agg gtg atc agc ctg aca gcc tgc ctg gtg aa  cag aac ctt cga ctg gac tgc cgt cat

                   E   N   N   T   N   L   P   I   Q   H   E   F   S   L   T   R   E   K   K   K
env     gaa aat aat aca cct ttg cca ata caa cat gaa ttt tca tta acg cgt gaa aaa aaa aaa
wt      gag aat aac acc aac ttg ccc atc cag cat gag ttc agc ctg acc cga gag aag aag aag

                   H   V   L   S   G   T   L   G   V   P   E   H   T   Y   R   S   R   V   N   L
env     cat gta tta agt gga aca tta gga gta cca gaa cat aca tat aga agt aga gta aat ttg
wt      cac gtg ctg tca ggc acc ctg ggg gtt ccc gag cac act tac cgc tcc cgc gtc aac ctt

                   F   S   D   R   F   I   K   V   L   T   L   A   N   F   T   T   K   D   E   G
env     ttt agt gat aga ttc ata aaa gta tta aca tta gca aat ttt aca aca aaa gat gaa gga
wt      ttc agt gac cgc ttt atc aag gtc ctt act cta gcc aac ttc acc acc aag gat gag ggc

                   D   Y   M   C   E   L   R   V   S   G   Q   N   P   T   S   S   N   K   T   I
env     gat tat atg tgt gag ctc aga gta agt gga caa aat cca aca agt agt aat aaa aca ata
wt      gac tac atg tgt gaa ctt cga gtc tcg ggc cag aat ccc aca agc tcc aat aaa act atc

                   N   V   I   R   D   K   L   V   K   C   G   G   I   S   L   L   V   Q   N   T
env     aat gta ata aga gat aaa tta gta aaa tgt gga gga ata agt tta tta gta caa aat aca
wt      aat gtg atc aga gac aag ctg gtc aag tgt ggt ggc ata agc ctg ctg gtt caa aac act

                   S   W   L   L   L   L   L   L   S   L   S   F   L   Q   A   T   D   F   I   S
env     agt tgg tta tta tta tta tta tta agt tta agt ttt tta caa gca aca gat ttt ata agt
wt      tcc tgg ctg ctg ctg ctc ctg ctt tcc ctc tcc ttc ctc caa gcc acg gac ttc att tct

                   L   *
env     tta tga
wt      ctg tga
```

FIGURE 6

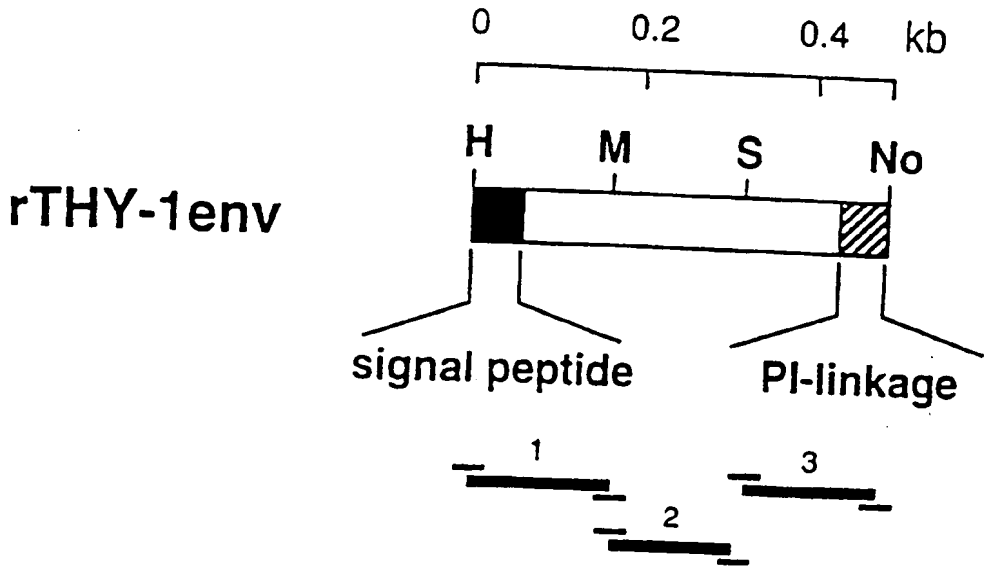


FIGURE 7

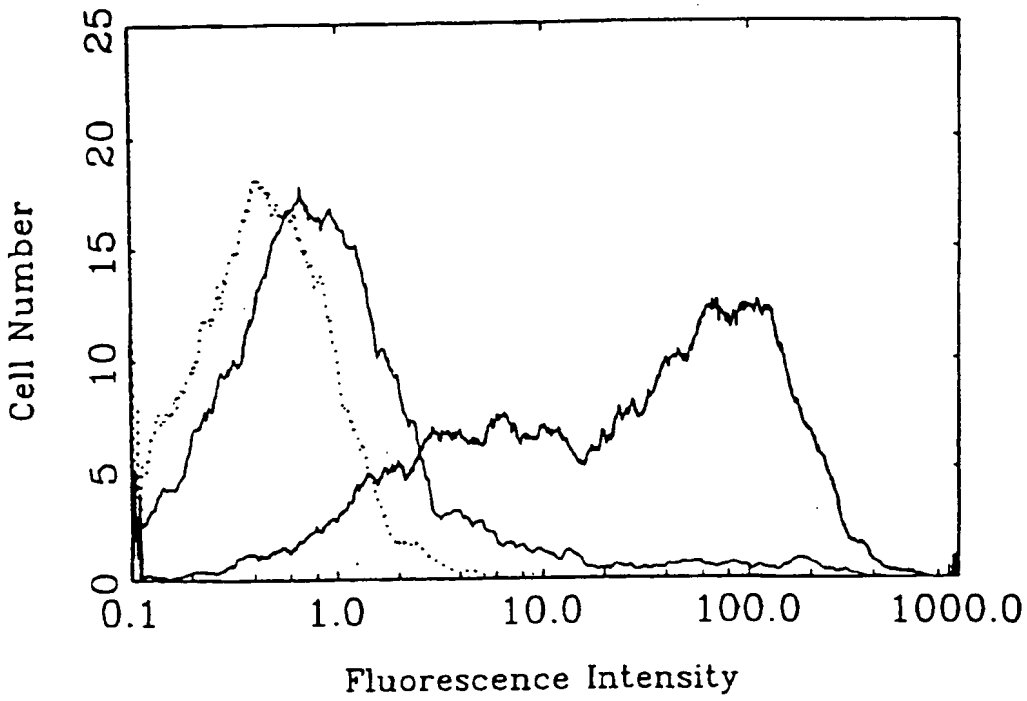

Cell Number
0
5
10
15
20
25
0.1
1.0
10.0
100.0
1000.0
Fluorescence Intensity

FIGURE 8

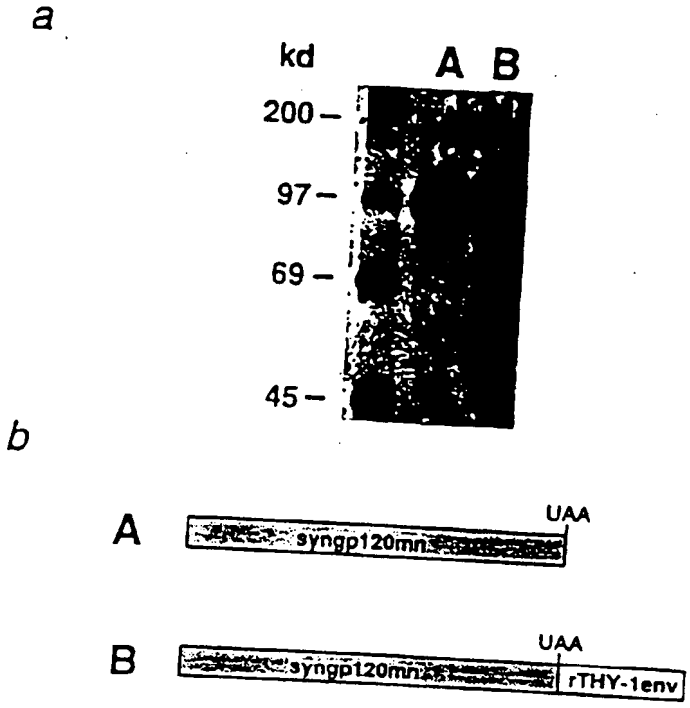
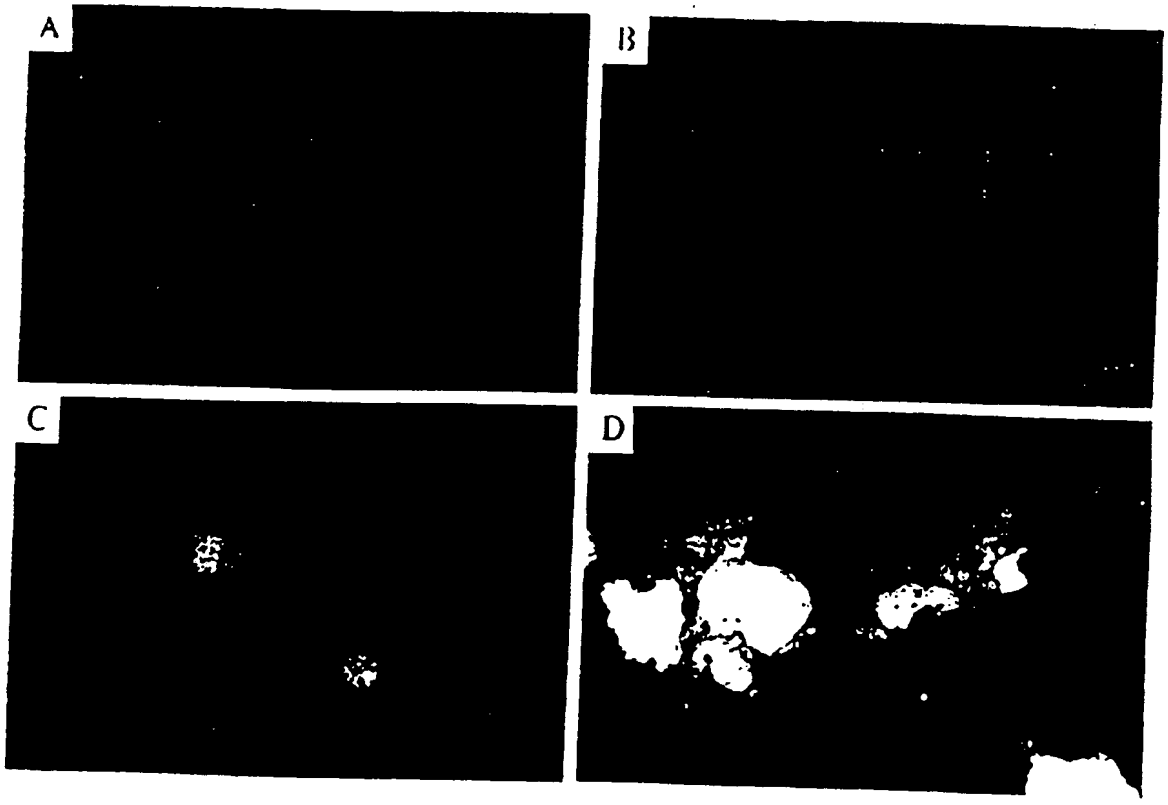

a
kd
A B
200
97
69
45
b
A
UAA
syngp120mn
B
UAA
syngp120mn
rTHY-1env

FIGURE 9

FIG. 10



```
   1  GAATTCACGC GTAAGCTTGC CGCCACCATG GTGAGCAAGG GCGAGGAGCT
  51  GTTCACCGGG GTGGTGCCCA TCCTGGTCGA GCTGGACGGC GACGTGAACG
 101  GCCACAAGTT CAGCGTGTCC GGCGAGGGCG AGGGCGATGC CACCTACGGC
 151  AAGCTGACCC TGAAGTTCAT CTGCACCACC GGCAAGCTGC CCGTGCCCTG
 201  GCCCACCCTC GTGACCACCT TCAGCTACGG CGTGCAGTGC TTCAGCCGCT
 251  ACCCCGACCA CATGAAGCAG CACGACTTCT TCAAGTCCGC CATGCCCGAA
 301  GGCTACGTCC AGGAGCGCAC CATCTTCTTC AAGGACGACG GCAACTACAA
 351  GACCCGCGCC GAGGTGAAGT TCGAGGGCGA CACCCTGGTG AACCGCATCG
 401  AGCTGAAGGG CATCGACTTC AAGGAGGACG GCAACATCCT GGGGCACAAG
 451  CTGGAGTACA ACTACAACAG CCACAACGTC TATATCATGG CCGACAAGCA
 501  GAAGAACGGC ATCAAGGTGA ACTTCAAGAT CCGCCACAAC ATCGAGGACG
 551  GCAGCGTGCA GCTCGCCGAC CACTACCAGC AGAACACCCC CATCGGCGAC
 601  GGCCCCGTGC TGCTGCCCGA CAACCACTAC CTGAGCACCC AGTCCGCCCT
 651  GAGCAAAGAC CCCAACGAGA AGCGCGATCA CATGGTCCTG CTGGAGTTCG
 701  TGACCGCCGC CGGGATCACT CACGGCATGG ACGAGCTGTA CAAGTAAAGC
 751  GGCCGCGGAT CC    (SEQ ID NO: 40)
```

FIG. 11

Native Factor VIII B domain deleted gene segment inserted in the expression vector

```
   1  AAGCTTAAAC CATGCCCATG GGGTCTCTGC AACCGCTGGC CACCTTGTAC
  51  CTGCTGGGGA TGCTGGTCGC TTCCGTGCTA GCCGCCACCA GAAGATACTA
 101  CCTGGGTGCA GTGGAACTGT CATGGGACTA TATGCAAAGT GATCTCGGTG
 151  AGCTGCCTGT GGACGCAAGA TTTCCTCCTA GAGTGCCAAA ATCTTTTCCA
 201  TTCAACACCT CAGTCGTGTA CAAAAAGACT CTGTTTGTAG AATTCACGGA
 251  TCACCTTTTC AACATCGCTA AGCCAAGGCC ACCCTGGATG GGTCTGCTAG
 301  GTCCTACCAT CCAGGCTGAG GTTTATGATA CAGTGGTCAT TACACTTAAG
 351  AACATGGCTT CCCATCCTGT CAGTCTTCAT GCTGTTGGTG TATCCTACTG
 401  GAAAGCTTCT GAGGGAGCTG AATATGATGA TCAGACCAGT CAAAGGGAGA
 451  AAGAAGATGA TAAAGTCTTC CCTGGTGGAA GCCATACATA TGTCTGGCAG
 501  GTCCTGAAAG AGAATGGTCC AATGGCCTCT GACCCACTGT GCCTTACCTA
 551  CTCATATCTT TCTCATGTGG ACCTGGTAAA AGACTTGAAT TCAGGCCTCA
 601  TTGGAGCCCT ACTAGTATGT AGAGAAGGGA GTCTGGCCAA GGAAAAGACA
 651  CAGACCTTGC ACAAATTTAT ACTACTTTTT GCTGTATTTG ATGAAGGGAA
 701  AAGTTGGCAC TCAGAAACAA AGAACTCCTT GATGCAGGAT AGGGATGCTG
 751  CATCTGCTCG GGCCTGGCCT AAAATGCACA CAGTCAATGG TTATGTAAAC
 801  AGGTCTCTGC CAGGTCTGAT TGGATGCCAC AGGAAATCAG TCTATTGGCA
 851  TGTGATTGGA ATGGGCACCA CTCCTGAAGT GCACTCAATA TTCCTCGAAG
 901  GTCACACATT TCTTGTGAGG AACCATCGCC AGGCGTCCTT GGAAATCTCG
 951  CCAATAACTT TCCTTACTGC TCAAACACTC TTGATGGACC TTGGACAGTT
1001  TCTACTGTTT TGTCATATCT CTTCCCACCA ACATGATGGC ATGGAAGCTT
1051  ATGTCAAAGT AGACAGCTGT CCAGAGGAAC CCCAACTACG AATGAAAAAT
1101  AATGAAGAAG CGGAAGACTA TGATGATGAT CTTACTGATT CTGAAATGGA
1151  TGTGGTCAGG TTTGATGATG ACAACTCTCC TTCCTTTATC CAAATTCGCT
1201  CAGTTGCCAA GAAGCATCCT AAAACTTGGG TACATTACAT TGCTGCTGAA
1251  GAGGAGGACT GGGACTATGC TCCCTTAGTC CTCGCCCCCG ATGACAGAAG
1301  TTATAAAAGT CAATATTTGA ACAATGGCCC TCAGCGGATT GGTAGGAAGT
1351  ACAAAAAAGT CCGATTTATG GCATACACAG ATGAAACCTT TAAGACTCGT
1401  GAAGCTATTC AGCATGAATC AGGAATCTTG GGACCTTTAC TTTATGGGGA
1451  AGTTGGAGAC ACACTGTTGA TTATATTTAA GAATCAAGCA AGCAGACCAT
1501  ATAACATCTA CCCTCACGGA ATCACTGATG TCCGTCCTTT GTATTCAAGG
1551  AGATTACCAA AAGGTGTAAA ACATTTGAAG GATTTTCCAA TTCTGCCAGG
1601  AGAAATATTC AAATATAAAT GGACAGTGAC TGTAGAAGAT GGGCCAACTA
1651  AATCAGATCC TCGGTGCCTG ACCCGCTATT ACTCTAGTTT CGTTAATATG
1701  GAGAGAGATC TAGCTTCAGG ACTCATTGGC CCTCTCCTCA TCTGCTACAA
1751  AGAATCTGTA GATCAAAGAG GAAACCAGAT AATGTCAGAC AAGAGGAATG
1801  TCATCCTGTT TTCTGTATTT GATGAGAACC GAAGCTGGTA CCTCACAGAG
1851  AATATACAAC GCTTTCTCCC CAATCCAGCT GGAGTGCAGC TTGAGGATCC
1901  AGAGTTCCAA GCCTCCAACA TCATGCACAG CATCAATGGC TATGTTTTTG
1951  ATAGTTTGCA GTTGTCAGTT TGTTTGCATG AGGTGGCATA CTGGTACATT
2001  CTAAGCATTG GAGCACAGAC TGACTTCCTT TCTGTCTTCT TCTCTGGATA
2051  TACCTTCAAA CACAAAATGG TCTATGAAGA CACACTCACC CTATTCCCAT
2101  TCTCAGGAGA AACTGTCTTC ATGTCGATGG AAAACCCAGG TCTATGGATT
2151  CTGGGGTGCC ACAACTCAGA CTTTCGGAAC AGAGGCATGA CCGCCTTACT
2201  GAAGGTTTCT AGTTGTGACA AGAACACTGG TGATTATTAC GAGGACAGTT
2251  ATGAAGATAT TTCAGCATAC TTGCTGAGTA AAAACAATGC CATTGAACCA
2301  AGAAGCTTCT CCCAGAATTC AAGACACCCT AGCACTAGGC AAAAGCAATT
2351  TAATGCCACC CCACCAGTCT TGAAACGCCA TCAACGGGAA ATAACTCGTA
2401  CTACTCTTCA GTCAGATCAA GAGGAAATTG ACTATGATGA TACCATATCA
2451  GTTGAAATGA AGAAGGAAGA TTTTGACATT TATGATGAGG ATGAAAATCA
2501  GAGCCCCCGC AGCTTTCAAA AGAAAACACG ACACTATTTT ATTGCTGCAG
2551  TGGAGAGGCT CTGGGATTAT GGGATGAGTA GCTCCCCACA TGTTCTAAGA
2601  AACAGGGCTC AGAGTGGCAG TGTCCCTCAG TTCAAGAAAG TTGTTTTCCA
2651  GGAATTTACT GATGGCTCCT TTACTCAGCC CTTATACCGT GGAGAACTAA
2701  ATGAACATTT GGGACTCCTG GGGCCATATA TAAGAGCAGA AGTTGAAGAT
```

Fig. 12

(1 of 2)

```
2751  AATATCATGG TAACTTTCAG AAATCAGGCC TCTCGTCCCT ATTCCTTCTA
2801  TTCTAGCCTT ATTTCTTATG AGGAAGATCA GAGGCAAGGA GCAGAACCTA
2851  GAAAAAACTT TGTCAAGCCT AATGAAACCA AAACTTACTT TTGGAAAGTG
2901  CAACATCATA TGGCACCCAC TAAAGATGAG TTTGACTGCA AAGCCTGGGC
2951  TTATTTCTCT GATGTTGACC TGGAAAAAGA TGTGCACTCA GGCCTGATTG
3001  GACCCCTTCT GGTCTGCCAC ACTAACACAC TGAACCCTGC TCATGGGAGA
3051  CAAGTGACAG TACAGGAATT TGCTCTGTTT TTCACCATCT TTGATGAGAC
3101  CAAAAGCTGG TACTTCACTG AAAATATGGA AAGAAACTGC AGGGCTCCCT
3151  GCAATATCCA GATGGAAGAT CCCACTTTTA AAGAGAATTA TCGCTTCCAT
3201  GCAATCAATG GCTACATAAT GGATACACTA CCTGGCTTAG TAATGGCTCA
3251  GGATCAAAGG ATTCGATGGT ATCTGCTCAG CATGGGCAGC AATGAAAACA
3301  TCCATTCTAT TCATTTCAGT GGACATGTGT TCACTGTACG AAAAAAAGAG
3351  GAGTATAAAA TGGCACTGTA CAATCTCTAT CCAGGTGTTT TTGAGACAGT
3401  GGAAATGTTA CCATCCAAAG CTGGAATTTG GCGGGTGGAA TGCCTTATTG
3451  GCGAGCATCT ACATGCTGGG ATGAGCACAC TTTTTCTGGT GTACAGCAAT
3501  AAGTGTCAGA CTCCCCTGGG AATGGCTTCT GGACACATTA GAGATTTTCA
3551  GATTACAGCT TCAGGACAAT ATGGACAGTG GGCCCCAAAG CTGGCCAGAC
3601  TTCATTATTC CGGATCAATC AATGCCTGGA GCACCAAGGA GCCCTTTTCT
3651  TGGATCAAGG TGGATCTGTT GGCACCAATG ATTATTCACG GCATCAAGAC
3701  CCAGGGTGCC CGTCAGAAGT TCTCCAGCCT CTACATCTCT CAGTTTATCA
3751  TCATGTATAG TCTTGATGGG AAGAAGTGGC AGACTTATCG AGGAAATTCC
3801  ACTGGAACCT TAATGGTCTT CTTTGGCAAT GTGGATTCAT CTGGGATAAA
3851  ACACAATATT TTTAACCCTC CAATTATTGC TCGATACATC CGTTTGCACC
3901  CAACTCATTA TAGCATTCGC AGCACTCTTC GCATGGAGTT GATGGGCTGT
3951  GATTTAAATA GTTGCAGCAT GCCATTGGGA ATGGAGAGTA AAGCAATATC
4001  AGATGCACAG ATTACTGCTT CATCCTACTT TACCAATATG TTTGCCACCT
4051  GGTCTCCTTC AAAAGCTCGA CTTCACCTCC AAGGGAGGAG TAATGCCTGG
4101  AGACCTCAGG TGAATAATCC AAAAGAGTGG CTGCAAGTGG ACTTCCAGAA
4151  GACAATGAAA GTCACAGGAG TAACTACTCA GGGAGTAAAA TCTCTGCTTA
4201  CCAGCATGTA TGTGAAGGAG TTCCTCATCT CCAGCAGTCA AGATGGCCAT
4251  CAGTGGACTC TCTTTTTTCA GAATGGCAAA GTAAAGTTTT TTCAGGGAAA
4301  TCAAGACTCC TTCACACCTG TGGTGAACTC TCTAGACCCA CCGTTACTGA
4351  CTCGCTACCT TCGAATTCAC CCCCAGAGTT GGGTGCACCA GATTGCCCTG
4401  AGGATGGAGG TTCTGGGCTG CGAGGCACAG GACCTCTACT GAGGGTGGCC
4451  ACTGCAGCAC CTGCCACTGC CGTCACCTCT CCCTCCTCAG CTCCAGGGCA
4501  GTGTCCCTCC CTGGCTTGCC TTCTACCTTT GTGCTAAATC CTAGCAGACA
4551  CTGCCTTGAA GCCTCCTGAA TTAACTATCA TCAGTCCTGC ATTTCTTTGG
4601  TGGGGGGCCA GGAGGGTGCA TCCAATTTAA CTTAACTCTT ACCGTCGACC
4651  TGCAGGCCCA ACGCGGCCGC
```

Fig. 12

(2 of 2)

Synthetic Factor VIII B domain deleted gene segment inserted in the expression vector

```
   1  AAGCTTAAAC CATGCCCATG GGGTCTCTGC AACCGCTGGC CACCTTGTAC
  51  CTGCTGGGGA TGCTGGTCGC TTCCGTGCTA GCCGCCACCC GCCGCTACTA
 101  CCTGGCCGCC GTGGAGCTGT CCTGGGACTA CATGCAGAGC GACCTGGGCG
 151  AGCTCCCCGT GGACGCCCGC TTCCCCCCCC GCGTGCCCAA GAGCTTCCCC
 201  TTCAACACCA GCGTGGTGTA CAAGAAAACC CTGTTCGTGG AGTTCACCGA
 251  CCACCTGTTC AACATTGCCA AGCCGCGCCC CCCCTGGATG GGCCTGCTGG
 301  GCCCCACCAT CCAGGCCGAG GTGTACGACA CCGTGGTGAT CACCCTGAAG
 351  AACATGGCCA GCCACCCCGT CAGCCTGCAC GCCGTGGGCG TGAGCTACTG
 401  GAAGGCCAGC GAGGGCGCCG AGTACGACGA CCAGACGTCC CAGCGCGAGA
 451  AGGAGGACGA CAAGGTGTTC CCGGGGGGGA GCCACACCTA CGTGTGGCAG
 501  GTGCTTAAGG AGAACGGCCC TATGGCCAGC GACCCCCTGT GCCTGACCTA
 551  CAGCTACCTG AGCCACGTGG ACCTGGTGAA GGATCTGAAC AGCGGGCTGA
 601  TCGGCGCCCT GCTGGTGTGT CGCGAGGGCA GCCTGGCCAA GGAGAAAACC
 651  CAGACCCTGC ACAAGTTCAT CCTGCTGTTC GCCGTGTTCG ACGAGGGGAA
 701  GAGCTGGCAC AGCGAGACTA AGAACAGCCT GATGCAGGAC CGCGACGCCG
 751  CCAGCGCCCG CGCCTGGCCC AAGATGCACA CCGTTAACGG CTACGTGAAC
 801  CGCAGCCTGC CCGGCCTGAT CGGCTGCCAC CGCAAGAGCG TGTACTGGCA
 851  CGTCATCGGC ATGGGCACCA CCCCTGAGGT GCACAGCATC TTCCTGGAGG
 901  GCCACACCTT CCTGGTGCGC AACCACCGCC AGGCCAGCCT GGAGATCAGC
 951  CCCATCACCT TCCTGACTGC CCAGACCCTG CTGATGGACC TAGGCCAGTT
1001  CCTGCTGTTC TGCCACATCA GCAGCCACCA GCACGACGGC ATGGAGGCTT
1051  ACGTGAAGGT GGACAGCTGC CCCGAGGAGC CCCAGCTGCG CATGAAGAAC
1101  AACGAGGAGG CCGAGGACTA CGACGACGAC CTGACCGACA GCGAGATGGA
1151  TGTCGTACGC TTCGACGACG ACAACAGCCC CAGCTTCATC CAGATCCGCA
1201  GCGTGGCCAA GAAGCACCCT AAGACCTGGG TGCACTACAT CGCCGCCGAG
1251  GAGGAGGACT GGGACTACGC CCCGCTAGTA CTGGCCCCCG ACGACCGCAG
1301  CTACAAGAGC CAGTACCTGA ACAACGGCCC CCAGCGCATC GGCCGCAAGT
1351  ACAAGAAGGT GCGCTTCATG GCCTACACCG ACGAGACTTT CAAGACCCGC
1401  GAGGCCATCC AGCACGAGTC CGGCATCCTC GGCCCCCTGC TGTACGGCGA
1451  GGTGGGCGAC ACCCTGCTGA TCATCTTCAA GAACCAGGCC AGCAGGCCCT
1501  ACAACATCTA CCCCCACGGC ATCACCGACG TGCGCCCCCT GTACAGCCGC
1551  CGCCTGCCCA AGGGCGTGAA GCACCTGAAG GACTTCCCCA TCCTGCCCGG
1601  CGAGATCTTC AAGTACAAGT GGACCGTGAC CGTGGAGGAC GGCCCCACCA
1651  AGAGCGACCC CCGCTGCCTG ACCCGCTACT ACAGCAGCTT CGTGAACATG
1701  GAGCGCGACC TGGCCTCCGG ACTGATCGGC CCCCTGCTGA TCTGCTACAA
1751  GGAGAGCGTG GACCAGCGCG GCAACCAGAT CATGAGCGAC AAGCGCAACG
1801  TGATCCTGTT CAGCGTGTTC GACGAGAACC GCAGCTGGTA TCTGACCGAG
1851  AACATCCAGC GCTTCCTGCC CAACCCCGCT GGCGTGCAGC TGGAAGATCC
1901  CGAGTTCCAG GCCAGCAACA TCATGCACAG CATCAACGGC TACGTGTTCG
1951  ACAGCCTGCA GCTGAGCGTG TGCCTGCATG AGGTGGCCTA CTGGTACATC
2001  CTGAGCATCG GCGCCCAGAC CGACTTCCTG AGCGTGTTCT TCTCCGGGTA
2051  TACCTTCAAG CACAAGATGG TGTACGAGGA CACCCTGACC CTGTTCCCCT
2101  TCTCCGGCGA GACTGTGTTC ATGTCTATGG AGAACCCCGG CCTGTGGATT
2151  CTGGGCTGCC ACAACAGCGA CTTCCGCAAC CGCGGCATGA CTGCCCTGCT
2201  GAAAGTCTCC AGCTGCGACA AGAACACCGG CGACTACTAC GAGGACAGCT
2251  ACGAGGACAT CTCCGCCTAC CTGCTGTCCA AGAACAACGC CATCGAGCCC
2301  CGCTCCTTCT CCCAAAACTC CCGCCACCCC AGCACGCGTC AGAAGCAGTT
2351  CAACGCCACC CCCCCCGTGC TGAAGCGCCA CCAGCGCGAG ATCACCCGCA
2401  CCACCCTGCA AAGCGACCAG GAGGAGATCG ACTACGACGA CACCATCAGC
2451  GTGGAGATGA AGAAGGAGGA CTTCGACATC TACGACGAGG ACGAGAACCA
2501  GAGCCCCCGC TCCTTCCAAA AGAAAACCCG CCACTACTTC ATCGCCGCCG
2551  TGGAGCGCCT GTGGGACTAC GGCATGAGCA GCAGCCCCCA CGTCCTGCGC
2601  AACCGCGCCC AGAGCGGCAG CGTGCCCCAG TTCAAGAAGG TGGTGTTCCA
2651  GGAGTTCACC GACGGCAGCT TCACCCAGCC CCTGTACCGC GGCGAGCTGA
2701  ACGAGCACCT GGGCCTGCTC GGCCCCTACA TCCGCGCCGA GGTGGAGGAC
```

Fig. 13

(1 of 2)

```
2751  AACATCATGG TGACCTTCCG CAACCAAGCC TCCCGGCCCT ACTCCTTCTA
2801  CTCCTCCCTG ATCAGCTACG AGGAGGACCA GCGCCAGGGC GCCGAGCCCC
2851  GCAAGAACTT CGTGAAGCCC AACGAGACTA AGACCTACTT CTGGAAGGTG
2901  CAGCACCACA TGGCCCCCAC CAAGGACGAG TTCGACTGCA AGGCCTGGGC
2951  CTACTTCAGC GACGTGGACC TGGAGAAGGA CGTGCACAGC GGCCTGATCG
3001  GCCCCCTGCT GGTGTGCCAC ACCAACACCC TGAACCCCCC CCACGGGAGG
3051  CAGGTGACTG TGCAGGAATT TGCCCTGTTC TTCACCATCT TCGACGAGAC
3101  TAAGAGCTGG TACTTCACCG AGAACATGGA GCGCAACTGC CGCGCCCCCT
3151  GCAACATCCA GATGGAAGAT CCCACCTTCA AGGAGAACTA CCGCTTCCAC
3201  GCCATCAACG GCTACATCAT GGACACCCTG CCCGGCCTGG TGATGGCCCA
3251  GGACCAGCGC ATCCGCTGGT ACCTGCTGTC TATGGGCAGC AACGAGAACA
3301  TCCACAGCAT CCACTTCAGC GGCCACGTTT TCACCGTGCG CAAGAAGGAG
3351  GAGTACAAGA TGGCCCTGTA CAACCTGTAC CCCGGCGTGT TCGAGACTGT
3401  GGAGATGCTG CCCAGCAAGG CCGGGATCTG GCGCGTGGAG TGCCTGATCG
3451  GCGAGCACCT GCACGCCGGC ATGAGCACCC TGTTCCTGGT GTACAGCAAC
3501  AAGTGCCAGA CCCCCCTGGG CATGGCCAGC GGCCACATCC GCGACTTCCA
3551  GATCACCGCC AGCGGCCAGT ACGGCCAGTG GGCTCCCAAG CTGGCCCGCC
3601  TGCACTACAG CGGCAGCATC AACGCCTGGT CGACCAAGGA GCCCTTCTCC
3651  TGGATCAAGG TGGACCTGCT GGCCCCCATG ATCATCCACG GCATCAAGAC
3701  CCAGGGCGCC CGCCAGAAGT TCAGCAGCCT GTACATCAGC CAGTTCATCA
3751  TCATGTACTC TCTAGACGGC AAGAAGTGGC AGACCTACCG CGGCAACAGC
3801  ACCGGCACCC TGATGGTGTT CTTCGGCAAC GTGGACAGCA GCGGCATCAA
3851  GCACAACATC TTCAACCCCC CCATCATCGC CCGCTACATC CGCCTGCACC
3901  CCACCCACTA CAGCATCCGC AGCACCCTGC GCATGGAGCT GATGGGCTGC
3951  GACCTGAACA GCTGCAGCAT GCCCCTGGGC ATGGAGAGCA AGGCCATCAG
4001  CGACGCCCAG ATCACCGCCT CCAGCTACTT CACCAACATG TTCGCCACCT
4051  GGAGCCCCAG CAAGGCCCGC CTGCACCTGC AGGGCCGCAG CAACGCCTGG
4101  CGCCCCCAGG TGAACAACCC CAAGGAGTGG CTGCAGGTGG ACTTCCAGAA
4151  AACCATGAAG GTGACTGGCG TGACCACCCA GGGCGTCAAG AGCCTGCTGA
4201  CCAGCATGTA CGTGAAGGAG TTCCTGATCA GCAGCAGCCA GGACGGCCAC
4251  CAGTGGACCC TGTTCTTCCA AAACGGCAAG GTGAAGGTGT TCCAGGGCAA
4301  CCAGGACAGC TTCACACCGG TCGTGAACAG CCTGGACCCC CCCCTGCTGA
4351  CCCGCTACCT GCGCATCCAC CCCCAGAGCT GGGTGCACCA GATCGCCCTG
4401  CGCATGGAGG TGCTGGGCTG CGAGGCCCAG GACCTGTACT GAAGCGGCCG
4451  C
```

Fig. 13

(2 of 2)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/16639

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C07H 21/04; C12P 21/02; C12N 15/11, 15/33, 15/48, 15/85
US CL :435/69.1, 70.1, 70.3, 172.3, 320.1; 536/23.1, 23.72, 25.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 70.1, 70.3, 172.3, 320.1; 536/23.1, 23.72, 25.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, EMBASE, MEDLINE, DERWENT
search terms: gene?, dna?, nucleic acid?, deoxyribonucleic?, synthe?, prefer? non-prefer? codon?

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96/09378 A (THE GENERAL HOSPITAL CORPORATION) 28 March 1996, abstract, page 1, line 20-page 4, line 26, page 15, lines 25-32, page 17, lines 27-39 and pages 42-54.	1-28
A	SEETHARAM et al. Mistranslation in IGF-1 During Over-Espression of the Protein in Escherichia coli Using a Synthetic Gene Containing Low Frequency Codons. Biochem. Biophys. Res. Comm. 30 August 1988. Vol. 155. No. 1. entire document.	1-28

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
21 NOVEMBER 1997	22 DEC 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer NANCY J. DEGEN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*